

[Counsel on signature page]

**UNITED STATES DISTRICT COURT**  
**NORTHERN DISTRICT OF CALIFORNIA**  
**SAN JOSE DIVISION**

*In re Google Generative AI Copyright  
Litigation,*

Master File Case No.: 5:23-cv-03440-EKL  
Consolidated with Case No.: 5:24-cv-02531-EKL

**[PROPOSED] ORDER RE: DISCOVERY OF  
ELECTRONICALLY STORED  
INFORMATION**

Judge: Hon. Eumi K. Lee

1           **1.       PURPOSE**

2           This Order will govern discovery of electronically stored information (“ESI”) in this case  
3 as a supplement to the Federal Rules of Civil Procedure, this Court’s Guidelines for the  
4 Discovery of Electronically Stored Information, and any other applicable orders and rules. The  
5 failure of this Protocol to address any issue is without prejudice to any position that a Party may  
6 take on that issue. To the extent that an issue is not addressed by a provision herein, the Federal  
7 Rules of Civil Procedure, the Northern District of California eDiscovery Guidelines, and other  
8 applicable Standing Orders and Guidelines of this Court control, unless the Parties agree  
9 otherwise or ask the Court to decide a particular issue.

10           **2.       COOPERATION**

11           The parties are aware of the importance the Court places on cooperation and commit to  
12 cooperate in good faith throughout the matter consistent with this Court’s Guidelines for the  
13 Discovery of ESI.

14           **3.       PROPORTIONALITY**

15           Parties are expected to use reasonable, good faith, and proportional efforts to preserve,  
16 identify, and produce relevant information consistent with Fed. R. Civ. P. 26(b)(1)<sup>1</sup>. This  
17 includes identifying appropriate limits to discovery, including limits on custodians, identification  
18 of relevant subject matter, time periods for discovery, and other parameters to reasonably limit  
19 and guide preservation and discovery issues.

20           **4.       LIAISON**

21           The parties have identified liaisons to each other who are and will be knowledgeable  
22 about and responsible for discussing their respective ESI. Each e-discovery liaison will be, or  
23 have access to those who are, knowledgeable about the technical aspects of e-discovery,  
24 including the location, nature, accessibility, format, collection, search methodologies, and  
25

26  
27           <sup>1</sup> Information can originate in any form, including ESI and paper, and is not limited to  
28 information created or stored electronically.

1 production of ESI in this matter. The parties will rely on the liaisons, as needed, to confer about  
 2 ESI and to help resolve disputes without court intervention.

3 a) Defendants appoint Paul Sampson and Kelly Knoll as their e-discovery  
 4 liaisons.

5 b) Plaintiffs appoint Yana Hart, Mark I. Richards, Greg Mullens, Christopher  
 6 Young, and Arielle Wagner as their e-discovery liaisons.

## 7 **5. PRESERVATION**

8 The parties have discussed their preservation obligations and needs and agree that  
 9 preservation of potentially relevant ESI will be reasonable and proportionate.

10 To reduce the costs and burdens of preservation and to ensure proper ESI is preserved,  
 11 the parties further agree that:

12 a) Only ESI created or received between July 11, 2018, and the present, will be  
 13 preserved.

14 b) The parties have discussed and will continue to meet and confer regarding the  
 15 types of ESI they believe should be preserved and the custodians, or general job titles or  
 16 descriptions of custodians, for whom they believe ESI should be preserved, *e.g.*, “HR head,”  
 17 “scientist,” and “marketing manager.” The parties shall add or remove custodians as reasonably  
 18 necessary.

19 c) The parties will endeavor to agree on the number of custodians per party for  
 20 whom ESI will be preserved, searched, reviewed, and produced.

21 d) The following data sources are not reasonably accessible because of undue  
 22 burden or cost pursuant to Fed. R. Civ. P. 26(b)(2)(B) and ESI from these sources will be  
 23 preserved pursuant to normal business retention, but not searched, reviewed, or produced:

- 24 1. backup systems and/or tapes used for disaster recovery; and
- 25 2. systems no longer in use that cannot be accessed.

e) Among the sources of data the parties agree are not reasonably accessible, or not proportional to the needs of the case, the parties agree not to preserve, search, review, or produce the following:

1. deleted, slack, fragmented, or other data accessible only by forensics;
2. random access memory (RAM), temporary files, or other ephemeral data that are difficult to preserve without disabling the operating system; and
3. on-line access data such as temporary internet files, history, cache, cookies, and the like.

## 6. SEARCH

a) In addition to the provisions agreed to below in this Section 6, the parties agree that in responding to an initial Fed. R. Civ. P. 34 request they will meet and confer about searching ESI in order to identify ESI that is subject to production in discovery, including the use (if any) of Technology Assisted Review (“TAR”) to filter out non-responsive documents, and a procedure for validating the efficacy of search terms. If any disagreements remain over the methods and protocols for searching ESI, including disputes regarding use of TAR or validation procedures, then the Parties agree to seek the Court’s resolution of their dispute.

b) The parties agree to filter out ESI that is not subject to discovery. Each party will use its best efforts to filter out common system files and application executable files by using a commercially reasonable hash identification process. Hash values that may be filtered out during this process are located in the National Software Reference Library (“NSRL”) NIST hash set list. Additional filtering of non-human readable files may include, but are not limited to, Bin files and P7S files.

c) A party is required to produce only a single copy of a responsive document, and the parties shall make reasonable efforts to de-duplicate ESI documents globally through an agreed upon metadata field, including MD5 Hash, Source Hash, or SHA256.

d) Email Threading: The Parties may use analytics technology to identify email threads and need only produce the unique most inclusive copy and related family members and may exclude lesser inclusive copies; EXCEPT that any thread member that contains additional or

1 altered data not contained in the most-inclusive email (including, *e.g.*, attachments or BCC  
2 recipients), is not a less inclusive email and must be separately produced. Responsibility for  
3 review and identification of potential less inclusive emails rests in the first instance with the  
4 Producing Party.

## 5 **7. PRODUCTION FORMATS**

6 Appendix A sets forth the types of metadata that will be included with each produced  
7 document, where available. In general, the Parties agree to produce documents, in TIFF file  
8 format. If particular documents warrant a different format, the Parties will cooperate to arrange  
9 for the mutually acceptable production of such documents. The Parties agree not to degrade the  
10 searchability of documents as part of the document production process. Parent-child relationships  
11 (the association between an attachment and its parent email, which are referred to as “families”)  
12 that have been maintained in the ordinary course of business must be preserved. The Producing  
13 Party will make reasonable efforts to produce responsive families together unless one or more  
14 members of the family are privileged, and in that case, the family will be Bates numbered with a  
15 placeholder for the withheld privileged document and the removed document will be logged on  
16 the Privilege Log. The Parties agree to meet and confer to address any issues related to  
17 production formats and processing.

18 The Parties also recognize that certain responsive information may reside in proprietary  
19 systems and formats or may consist of structured data. To the extent such information is to be  
20 produced in this case, the Parties agree to meet and confer about the form in which that  
21 information is produced.

## 22 **8. PHASING**

23 When a party propounds discovery requests pursuant to Fed. R. Civ. P. 34, the parties  
24 will discuss and agree to meet and confer about whether the production of ESI should be  
25 prioritized and how that prioritization may occur across ESI sources and custodians.  
26  
27  
28

1           **9.       HYPERLINKS**

2           Document(s) and/or folder(s) of documents that are hyperlinked inside a responsive  
3 document need not be produced in the first instance. The Requesting Party may request up to  
4 500 specifically identified hyperlinks in produced documents by identifying the Bates number  
5 and URL or link text for each requested link to the Producing Party, and the Producing Party  
6 will produce any responsive, non-privileged, point-in-time hyperlinked document. For each  
7 document produced under this Paragraph, the Producing Party shall include the following  
8 metadata fields, if available: Collaborators (accounts and groups that have permission to edit the  
9 file or add comments) and Viewers (accounts and groups that have permission to view the file).

10           **10.     PRIVILEGE LOG**

11           a)       If a party reasonably determines that one or more responsive documents are not  
12 discoverable because they are subject to the attorney-client communication privilege or work  
13 product doctrine, or another recognized protection or privilege (collectively, the “Privileges” and  
14 each a “Privilege”), the party shall produce a log for each document or each category of  
15 documents withheld for privilege from that production pursuant to the terms and parameters  
16 below. The privilege log(s) will be produced in MS Excel format or any other format that  
17 permits electronic sorting and searching.

18           b)       A producing party will produce an initial privilege log forty-five (45) days after  
19 its first production of documents in this action and subsequent privilege logs every forty-five  
20 (45) days during discovery.

21           c)       The Producing Party must provide a privilege log for all documents produced to  
22 date no later than thirty (30) days prior to the close of fact discovery, to allow the Parties  
23 adequate time to meet and confer to resolve any privilege disputes or seek resolution within  
24 seven (7) days after the fact discovery cut-off, consistent with Local Rule 37-3.

25           c)       For those documents or categories of similarly situated documents required to be  
26 placed on a privilege log pursuant to the Protective Order and ESI Order in this action, to the  
27 extent applicable, each Party’s privilege log must provide objective metadata (to the extent it is  
28

1 reasonably available and does not reflect privileged or protected information itself), an  
2 indication of the privilege or protection being asserted, identification of any associated persons  
3 who are attorneys, and a description of the privileged or protected material.

4 d) Objective metadata includes the following:

- 5 1. Custodian
- 6 2. Author
- 7 3. From
- 8 4. To
- 9 5. CC
- 10 6. BCC
- 11 7. Date Sent
- 12 8. Time Sent
- 13 9. Date Created
- 14 10. A unique identifying number for each logged document ad seriatim starting  
15 with the number 1 (the “Privilege Log ID” or “Item No.”); and
- 16 11. A family relationship identifier

17  
18 e) In addition to the objective metadata fields set for above, the Parties agree to  
19 include the following information on a privilege log, insofar as that information is reasonably  
20 available:

- 21 1. the Bates number range;
- 22 2. If the sender or recipient of an email communication is shown in the  
23 metadata on the log as a distribution list, the party asserting the claim of privilege  
24 shall provide the current membership of the list, unless the list has already been  
25  
26  
27  
28

deleted or is no longer populated (for example, if all employees that had been on the list have left employment with Defendants);<sup>2</sup>

3. the subject matter addressed in the document;

4. the purpose for which the information was prepared or communicated;

5. the nature of the privilege or protection asserted (*i.e.*, attorney-client privilege, work product doctrine);

6. an indication of whether the document has been produced in redacted form or withheld in its entirety; and

7. the specific basis for the claim that the information is privileged.

f) Communications involving counsel, or experts retained for the purpose of this litigation, that post-date the filing of the complaint on July 11, 2023 need not be placed on a privilege log.

g) Email chains: If there is more than one branch of (*i.e.*, more than one unique group of recipients of) an email thread, each branch will be individually logged; however, each individual email within the thread need not be logged if the recipients of the email chain are all identical or if a Party has elected to use threading for review and/or production of emails. A Party asserting privilege over a chain of emails may produce only a single redacted copy of such email chain to the extent some portions are only partially privileged, except that any unique branches of the email chain must also either be produced in redacted form or included on the metadata privilege log.

h) Privilege logs will be produced in an Excel format, or other agreed-upon format, which allows the receiving party to search and sort any and all columns and entries of the privilege log.

---

<sup>2</sup> Information concerning individuals' titles and/or job positions shall be provided in a separate list appended to the privilege log. Attorneys or their staff shall be identified in any privilege log with a unique identifier (*e.g.*, asterisk or double asterisk).



1           **11.     VALIDATION OF SEARCH TERMS**

2           After applying search terms to targeted document set(s), the Producing Party will review  
3 a statistically valid (68%,  $\pm 5$ ; one standard deviation), randomly generated sample set of the  
4 documents that do not hit upon any search terms (“Null Set Sample”); and produce to the  
5 Requesting Party responsive, non-privileged documents from that Null Set Sample. This  
6 production shall be clearly identified as the product of the Producing Party’s review of a Null  
7 Set Sample.

8           Based upon the responsive documents located within the Null Set Sample, the Parties will  
9 work cooperatively to develop additional, new search terms in the same manner used to develop  
10 the initial set of search terms. The Producing Party will apply these new terms to the targeted  
11 document set(s) (having already run the initial search terms) and produce responsive, non-  
12 privileged documents.

13           A one-time check of a Null Set Sample as described above provides quality control  
14 proportional to the needs of the litigation. Either Party may apply to the Court, following robust  
15 meet and confer, for adjustment to this process upon a showing of good cause founded upon  
16 unforeseen circumstances.

17           **12.     MODIFICATION**

18           This Stipulated Order may be modified by a Stipulated Order of the parties or by the  
19 Court for good cause shown. Any such modified Stipulated Order will be titled sequentially as  
20 follows, “First Modified Stipulated Order re: Discovery of Electronically Stored Information for  
21 Standard Litigation,” and each modified Stipulated Order will supersede the previous Stipulated  
22 Order.

23           **IT IS SO STIPULATED**, through Counsel of Record.

24  
25  
26  
27  
28

1 Dated: April 11, 2025

Respectfully submitted,

2 By: /s/ Lesley E. Weaver

3 Lesley E. Weaver (State Bar No. 191305)

4 Anne K. Davis (State Bar No. 267909)

5 Joshua D. Samra (State Bar No. 313050)

**BLEICHMAR FONTI & AULD LLP**

1330 Broadway, Suite 630

Oakland, CA 94612

Telephone: (415) 445-4003

Email: lweaver@bfalaw.com

adavis@bfalaw.com

jsamra@bfalaw.com

6 Gregory S. Mullens (*pro hac vice*)

**BLEICHMAR FONTI & AULD LLP**

75 Virginia Road, 2<sup>nd</sup> Floor

White Plains, NY 10603

Telephone: (415) 445-4006

Email: gmullens@bfalaw.com

13 /s/ Joseph R. Saveri

14 Joseph R. Saveri (State Bar No. 130064)

Cadio Zirpoli (State Bar No. 179108)

15 Christopher K.L. Young (State Bar No. 318371)

16 Elissa A. Buchanan (State Bar No. 249996)

Evan Creutz (State Bar No. 349728)

**JOSEPH SAVERI LAW FIRM, LLP**

601 California Street, Suite 1505

San Francisco, California 94108

Telephone: (415) 500-6800

Facsimile: (415) 395-9940

Email: jsaveri@saverilawfirm.com

czirpoli@saverilawfirm.com

cyoung@saverilawfirm.com

eabuchanan@saverilawfirm.com

ecreutz@saverilawfirm.com

23 Matthew Butterick (State Bar No. 250953)

1920 Hillhurst Avenue, #406

Los Angeles, CA 90027

Telephone: (323) 968-2632

Facsimile: (415) 395-9940

Email: mb@buttericklaw.com

/s/ Ryan J. Clarkson

Ryan J. Clarkson (State Bar No. 257074)

Yana Hart (State Bar No. 306499)

Mark I. Richards (State Bar No. 321252)

Tiara Aveness (State Bar No. 343928)

**CLARKSON LAW FIRM, P.C.**

22525 Pacific Coast Highway

Malibu, CA 90265

Telephone: (213) 788-4050

Email: rclarkson@clarksonlawfirm.com

yhart@clarksonlawfirm.com

mrichards@clarksonlawfirm.com

tavaness@clarksonlawfirm.com

Tracey Cowan (State Bar No. 250053)

**CLARKSON LAW FIRM, P.C.**

95 Third Street, Second Floor

San Francisco, CA 94103

Telephone (213) 788-4050

Email: tcowan@clarksonlawfirm.com

/s/ Stephen J. Teti

Stephen J. Teti (*pro hac vice*)

**LOCKRIDGE GRINDAL NAUEN PLLP**

265 Franklin St., Suite 1702

Boston, MA 02110

Tel.: (617) 456-7701

Email: sjteti@locklaw.com

Brian D. Clark (*pro hac vice*)

Laura M. Matson (*pro hac vice*)

Arielle Wagner (*pro hac vice*)

**LOCKRIDGE GRINDAL NAUEN PLLP**

100 Washington Avenue South, Suite 2200

Minneapolis, MN 55401

Telephone: 612)339-6900

Facsimile: (612)339-0981

Email: bdclark@locklaw.com

lmmatson@locklaw.com

aswagner@locklaw.com

*Counsel for Individual and Representative  
Plaintiffs and the Proposed Class*

1 Dated: April 11, 2025

/s/ Paul J. Sampson

David H. Kramer, SBN 168452

Email: dkramer@wsgr.com

Maura L. Rees, SBN 191698

Email: mrees@wsgr.com

**WILSON SONSINI GOODRICH & ROSATI  
Professional Corporation**

650 Page Mill Road

Palo Alto, CA 94304-1050

Telephone: (650) 493-9300

Eric P. Tuttle, SBN 248440

Email: eric.tuttle@wsgr.com

**WILSON SONSINI GOODRICH & ROSATI  
Professional Corporation**

701 Fifth Avenue, Suite 5100

Seattle, WA 98104-7036

Telephone: (206) 883-2500

Paul J. Sampson (*pro hac vice*)

Email: psampson@wsgr.com

**WILSON SONSINI GOODRICH & ROSATI  
Professional Corporation**

95 South State Street, Suite 1000

Salt Lake City, UT 84111

Telephone: (801) 401-8510

*Counsel for Defendants*

GOOGLE LLC AND ALPHABET INC.

**SIGNATURE ATTESTATION**

I, Lesley E. Weaver, attest that concurrence in the filing of this document has been  
obtained from the other signatories. I declare under penalty of perjury that the foregoing is true  
and correct.

Executed this 11<sup>th</sup> day of April, 2025, at Oakland, California.

By: /s/ Lesley E. Weaver

Lesley E. Weaver

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

**IT IS ORDERED** that the foregoing Agreement is approved.

Dated:

---

HON. SUSAN VAN KEULEN  
UNITED STATES MAGISTRATE JUDGE

## APPENDIX A

Field Name	Field Description
BEGBATES	Beginning Bates number as stamped on the production image
ENDBATES	Ending Bates number as stamped on the production image
BEGATTACH	First production Bates number of the first document in a family
ENDATTACH	Last production Bates number of the last document in a family
CUSTODIAN	Includes the Individual (Custodian) from whom the documents originated and all Individual(s) whose documents de-duplicated out (De-Duped Custodian)
HASH	MD5 or SHA-256 Hash Value
ATTACHNAME	The file name(s) of the attachment(s) to a parent document. Separated by semicolon
ATTACHCOUNT	Number of attachments to an email or embedded parent, as generated by commercially available discovery processing tools
SUBJECT	Subject line of email
TITLE	Title from properties of document
DATESENT	Date email was sent (format: MM/DD/YYYY)
DATERECEIVED	Date received and time (format: MM/DD/YYYY HH:MM)
TIMESENT	Time email was sent
TIMERECEIVED	Time email was received
TIMEZONE	Time zone used to process data during document collection and processing.
TO	All recipients that were included on the “To” line of the message
FROM	The name and email address of the sender of the message
CC	All recipients that were included on the “CC” line of the message
BCC	All recipients that were included on the “BCC” line of the message

AUTHOR	Original composer of document or sender of message
FILENAME	Filename of an electronic document (Edoc or attachment)
FILEEXTENSION	Indicates file extension of source of native file ( <i>e.g.</i> , .msg, .doc., .xls, etc.)
TEXTPATH	Relative path to the document level text file
PAGECOUNT	The number of pages of the Document, excluding the pages of documents in the same family
CONFIDENTIALITY	Identifies the confidential designation as applied in accordance with protective order on file in this Action
DATEMOD	Date an electronic document was last modified or created (format: MM/DD/YYYY) (Edoc or attachment)
DATECREATED	Date the document was created (format: MM/DD/YYYY) (Edoc or attachment)
NATIVELINK	Native File Link (Native Files only)
REDACTION	Populated “Yes” when a document contains a redaction
WITHHELD PLACEHOLDER	To the extent a document is fully withheld (on the basis of privilege or otherwise), this field must be populated with a “Y”